T^*GreenHDFS: A Cyber-Physical, Data-Centric Cooling Energy Costs Reduction Approach for Big Data Analytics Cloud

Rini T. Kaushik, Tarek Abdelzaher, Klara Nahrstedt
University of Illinois, Urbana-Champaign
{*kaushik1, zaher, klara*}*@illinois.edu*

Abstract

Big Data explosion and surge in large-scale Big Data analytics cloud infrastructure have led to burgeoning energy costs and present a challenge to the existing run-time cooling energy management techniques. T^*GreenHDFS, a thermal-aware cloud file system, takes a novel, data-centric approach to reduce cooling energy costs. On the physical-side, T^*GreenHDFS is cognizant of the uneven thermal-profile in the data centers due to complex airflow patterns, varying ability of the cooling system to cool different parts of the data center, and run-time load distribution. On the cyber-side, T^*GreenHDFS is aware of the differences in the data-semantics of the data placed on the clusters. Based on this knowledge, and coupled with its predictive data models, T^*GreenHDFS does proactive, thermal-aware data placement, which implicitly results in thermal-aware computation placement in the Big Data analytics cloud compute model. Evaluation results show up to 59% reduction in the cooling energy costs with T^*GreenHDFS.

1 Introduction

Explosion in Big Data [9] has led to a surge in *data-intensive* computing [13]. Data-intensive computing has myriad use cases such as fraud detection in financial transactions, building indexes and search rankings for the internet-scale search engines, log processing, mail anti-spam detection, click-stream processing, and machine learning and data mining on massive web logs to build predictive user interest models for advertising optimizations. Big Data storage and analysis necessitate infrastructure like cloud that allows massive scale-out at economies of scale.

Gartner predicts that by 2015, the data centers dedicated to cloud computing will account for 71 percent of worldwide data center hardware spending. The huge infrastructure brings in its wake burgeoning energy costs [16]. Energy consumed for computation and cooling is the dominant factor in data center run-time costs [3, 17]. A study of 22 data centers found average power usage efficiency (PUE) [2] value of 2 [18], which means that cooling energy costs (dominant part of the non-IT costs) can amount to almost half of the total energy costs.

The data centers that power the Big Data analytics cloud are different from the traditional data centers [19]; cost-efficiency is a very important metric given the sheer scale-out needed. Instead of using high-end, expensive components such as network switches with very high bisection bandwidth, these data centers use low-cost, commodity hardware. Because of the network bandwidth constraints of the commodity network switches and the huge data size of Big Data, sending data to the computations is no longer feasible; *data-locality* is an important consideration for high performance, and computations are sent to the servers where the data is residing. For example, MapReduce, a highly scalable, parallel processing framework widely used in Big Data analytics clouds, owes its high performance to its data-locality feature [14].

Majority of the existing research on run-time reduction of cooling energy costs relies on thermal-aware computation placement or migration to reduce the cooling energy costs [6, 41, 32, 31, 37, 4, 39, 35, 30]. These techniques are *data-placement agnostic* and attempt to place computationally heavy jobs on servers in a thermal-aware manner. These techniques work very well when servers are state-less, data resides on a shared SAN or NAS device, and data can be sent to the computation without network bandwidth constraints. Big Data analytics cloud has a different compute model and presents a challenge to the existing run-time cooling techniques.

In interest of performance, Big Data analytics' data-locality constraint restricts the server options in thermal-aware computation placement techniques to only the servers that host a replica of the data to be computed upon; thereby, reducing the potential cooling energy savings. On the other hand, neglecting data-locality results in higher cooling energy savings at the cost of performance. Other cooling management techniques use computation migration; they reactively migrate computations from a server with high run-time temperature to lower temperature servers. Computation migration is viable only when servers are state-less; in Big Data analytics cloud servers have significant state. In addition, computation migration to a server that doesn't host a replica of the data results in non-local data accesses, which comes at a performance cost.

Big Data mandates data to be the first-class object in computing; data needs to be placed first in a thermal-aware manner at the creation-time itself, so the run-time scheduling of the computation jobs on to the servers hosting the data can enjoy cooling energy savings without compromising on data-local performance. The contribution of our paper is in the form of a novel, *data-centric* approach for reducing *cooling* energy costs in *Big Data analytics* cloud as shown in Figure 1. To the best of our knowledge, this is the *first paper* that takes a data-centric approach to reduce cooling energy costs in Big Data analytics cloud. In our earlier work, GreenHDFS focused only on reducing compute energy costs via a data-centric, purely cyber-side scale-down approach and had no thermal-awareness of the physical-side [23, 24].

Uneven thermal-profile exists in the data centers due to complex airflow patterns, varying ability of the cooling system to cool different parts of the data center, and run-time load distribution. T^*GreenHDFS combines its predictive data-semantics knowledge on the cyber-side with the *thermal-profile* knowledge of the cluster on the *physical-side* to do *proactive cyber-side thermal-aware data placement*. Thermal-aware data placement is not restricted in its server choices unlike the thermal-aware computation placement techniques discussed earlier. Since, computations are sent to the data in the Big Data cloud compute model, thermal-aware data placement inherently results in thermal-aware computation placement.

Evaluation with distributions from real-world traces at production Hadoop cluster at Yahoo! shows that T^*GreenHDFS significantly lowers cooling costs as it results in lower overall cluster temperature, more uniform thermal-profile, less thermal hotspots, and higher cooling efficiency of the cooling system compared to a baseline cluster with no thermal management. The remainder of the paper is structured as follows. In Section 2 and 3, we discuss the design and architecture of T^*GreenHDFS. In Section 4, we include experimental results demonstrating the effectiveness and robustness of our design and algorithms. In Section 5, we discuss related work and conclude.

2 Thermal-Aware T^*GreenHDFS

T^*GreenHDFS, a thermal-aware cloud file system, takes a novel data-centric cooling energy management approach. T^*GreenHDFS is proactively cognizant of the difference in the data semantics of the data that is to be placed on the cluster (cyber-side), and the uneven thermal-profile of the servers in the cluster (physical-side). T^*GreenHDFS adds thermal-aware mechanisms and associated states in data placement in the Hadoop Distributed File System (HDFS) [25], and assumes same replication, file chunking, fault-tolerance, and reliability mechanisms as the baseline HDFS and same MapReduce job scheduling policies and algorithms as Hadoop [44].

Uneven *thermal-profile* in data centers is because of complex airflow patterns, varying ability of the cooling system to cool different parts of the data center, and run-time computational load distribution. Just like the servers in the cluster vary in their run-time thermal-profile and cooling-efficiency, data in the cluster *differs* in semantics such as access-profiles, sizes, and lifespans; several data classes coexist in the compute clusters. Some data classes are heavily computed upon, and thereby have a high access-profile. Other data classes receive medium, low, or very low computations, and thereby have an accordingly medium, low or very-low access-profile. Yet another class of data just lies dormant in the cluster (i.e., without receiving any computations or accesses). Data also differs in the length and distribution of accesses over its lifespan. T^*GreenHDFS uses the following two thermal-aware mechanisms to save cooling energy costs.

Thermal-Aware Data Placement: T^*GreenHDFS maintains run-time thermal-profile of the servers in the *Hot* zone to do a *fine-grained* thermal-aware data placement of the files onto the servers resulting in a more uniform run-time thermal profile in the cluster. The cyber-controller gets run-time thermal-map of the servers every heartbeat from the physical-system covered in Section 3.1.1. When a file create request is received by the cyber-controller, it uses its predictive data models to predict the file's access-profile (translates to computation-profile). Based on the predicted access-profile, and run-time thermal-map, cyber-actuator does *thermal-aware data placement* as shown in Section 3.2.3; a file with high anticipated computation-profile is placed proactively on low run-time temperature server.

Thermal-Aware Server Zone Partitioning: The initial bootstrapping phase in the physical-system does a thermal-aware zone partitioning of the servers in the cluster into cost, power, performance and cooling-efficiency differentiated *Hot* and *Cold* data zones as shown in the Section 3.1.2. The *Cold* zone servers are used to store dormant data class, i.e., data class with very low or negligible data access-profile (cyber), and thereby low computation-profile, generating low-server energy, hence requiring low cooling energy. The *Cold* zone servers experience significant idleness and can be scaled-down, resulting in lower server operating costs as well as we have shown in our earlier paper [24]. The *Hot* data zone servers are used to store the rest of the data classes with high/medium/low data access-profile, generating high/medium-server energy, and hence requiring high/medium cooling energy.

T^*GreenHDFS assigns the most inherently cooling-inefficient servers in the cluster to the *Cold* zone. Thus, T^*GreenHDFS ensures that the cooling-inefficient servers receive negligible computations, don't generate much heat and their exhaust temperature remain bounded. Such a thermal-aware data zone partitioning reduces the hot spots in the cluster leading to an overall lower temperature in the cluster which in turn reduces the cooling energy costs. A migration policy running in the cyber-controller moves the data that has become dormant in the *Hot* zone to the *Cold* zone on a daily basis. The dormant data class can be surprisingly huge. Based on our analysis of a production Hadoop cluster at Yahoo!, we realized that 56% of the data in the cluster was dormant [24]. Other studies have found similar numbers. This data couldn't just be deleted as it was being stored for regulatory and com-

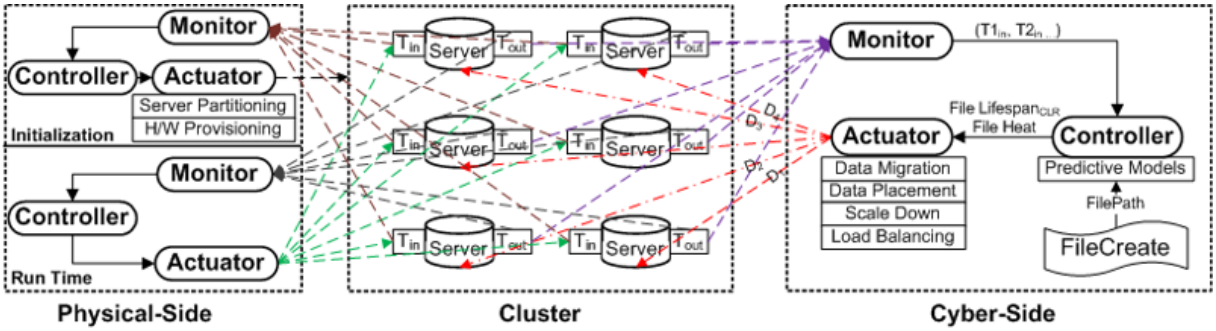

Figure 1: **Thermal-Aware T^*GreenHDFS.**

pliance purposes.

The thermal-aware data placement can be used in a standalone fashion cluster-wide if zoning and data migration are not feasible for an enterprise. In the evaluation, we cover several configurations of T^*GreenHDFS and show that proactive, thermal-aware, fine-grained data placement in itself results in significant cooling costs savings. However, thermal-aware data placement used in conjuction with thermal-aware zone partitioning leads to the highest possible cooling costs savings. In the next section, we cover the architecture of the cyber-physical system in T^*GreenHDFS.

3 T^*GreenHDFS Architecture

In this section, we cover the details of the physical-system component, and the physical- and cyber-side monitor, controller and actuator components in the cyber-physical system shown in the Figure 1 in T^*GreenHDFS. The cyber-side gets run-time thermal feedback from the physical-system to guide its thermal-aware data placement policies, which result in thermal changes in the physical-system. The physical-side monitors the run-time changes in the overall thermal temperature of the physical-system and makes changes to the CRAC setting which in turn changes the physical-system thermal profile.

3.1 Physical System

The physical-system comprises of the n servers and temperature sensors in the cluster. Each server in the cluster has a temperature sensor T_{Inlet} at its inlet and $T_{Exhaust}$ at its exhaust. In addition, there is a temperature sensor $T_{Overall}$ that measures the overall temperature in the cluster. The physical-system has two operating phases: 1) run-time phase, and 2) initial bootstrapping phase.

3.1.1 Run-Time Phase

During the run-time, the temperature sensors on all the n servers in the cluster monitor the exhaust temperature $T_{Exhaust}$ and pass the server thermal-map $T_{Exhaust_1}$, $T_{Exhaust_2}$,, $T_{Exhaust_n}$ to the cyber-monitor; which makes the cyber-controller aware of the run-time changes in the exhaust temperatures of the servers. The $T_{Overall}$ is also monitored and sent by the physical-monitor to the physical-controller; making it aware of the overall temperature in the cluster. In addition, cyber-monitor collects information about the free capacity available (capacity map), and utilization of the processors, disks, and network (utilization map) about each and every server and sends this information to the cyber-controller. Cyber-actuator uses the run-time information about the server exhaust temperatures to guide its fine-tuned thermal-aware data placement as covered in Section 3.2.3. Physical-controller uses the $T_{Overall}$ value to make decisions about the CRAC temperature as covered in Section 3.3. The monitored information is piggybacked on the heartbeat mechanism which is always in place in large-scale distributed file systems [25], to ensure that there is no additional performance overhead of monitoring.

3.1.2 Bootstrapping Phase

Even under uniform load, inlet temperature of the servers in the cluster inherently differs in a location-sensitive manner, and results in an uneven inlet thermal-profile in the cluster. The higher the inlet temperature of a server, the lower is its cooling-efficiency (i.e., ability to remove the heat generated); as a result, there are inherent variations in the cooling-efficiency of the servers in the cluster. This variation is due to the complex nature of airflow inside data centers; even when all the servers are uniformly loaded, some of the hot air from the outlets of the servers recirculates into the inlets of other servers. The recirculated hot air mixes with the supplied cold air and causes inlets of some of the servers in the data center to experience a rise in inlet temperature, thereby resulting in an *uneven* inlet thermal-profile. In addition, the CRACs vary in their ability to cool different places in a data center (e.g., a corner of the room, farthest from the CRAC), and further aid in the uneven inlet thermal-profile [6]. The uneven inlet thermal-profile (hence, cooling-efficiency) leads to an uneven exhaust thermal-profile. The most cooling-inefficient servers have much higher exhaust temperatures; development of such hot spots leads to an overall higher temperature in the cluster and results in higher cooling costs. Figure 2, shows a histogram of the range of the inlet temperatures present in the cluster.

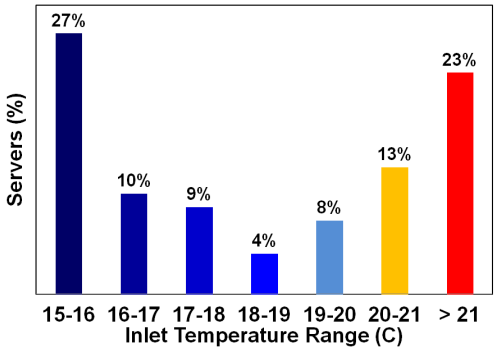

Figure 2: **The Histogram of the Inlet Temperatures in a Big Data Analytics cluster with no Energy Management.**

The initial bootstrapping phase is run prior to provisioning the servers in the cluster to: 1) rank the cooling-efficiencies of the servers, 2) partition the servers in the cluster into cooling-efficiency differentiated data zones, and 3) provision the hardware on the servers in a cost- and power-aware manner. The bootstrapping phase creates a thermal-profile of the cluster, with all servers kept at same utilization, to identify the inherently *cooling-inefficient servers*. The bootstrapping phase uses MentorGraphics floVENT [1], computational Fluid Dynamics (CFD) [34] simulator, to simulate the cluster under consideration.[1]

At the conclusion of the simulation, floVENT provides the inlet and the exhaust temperature for each server and the CRACs in the cluster. The bootstrapping phase then ranks the servers in the cluster in decreasing order of their inlet temperatures (i.e., cooling-efficiency). It also creates a per-rack ranking of the inlet temperature (i.e., cooling-efficiency) of the servers in the racks. The bootstrapping phase needs to be rerun if there are significant physical changes to the hardware or layout in the data center. Past research has shown that changes in the physical layout of the data center can result in significant changes in the air flow in the data center and hence, result in a change in the thermal profile of the servers [33].

Thermal-Aware Zone Partitioning The Thermal-Aware Zone Partitioning uses the thermal-efficiency ranking created in the earlier section to partition the servers in the cluster into cooling-efficiency differentiated zones. T^*GreenHDFS comes up with a number of k servers to be assigned to the *Cold* zone using a cyber-physical server provisioning algorithm (beyond the scope of this paper).[2] In the evaluation, we consider four values of k = 0, 10, 20, 30 in zone partitioning in T^*GreenHDFS. The 0 value corresponds to T^*GreenHDFS which treats the entire cluster the same, and doesn't do any zoning. We consider two options to assign servers to the zones in this paper:

Cluster-Level Zone Partitioning In this scheme, T^*GreenHDFS assigns the k number of most cooling-inefficient servers in the cluster to the *Cold* zone, and the remainder of the servers to the *Hot* zone. Typically the racks in the center rows of the data center have higher inlet temperatures than the racks in the outer-most rows. Hence, the central racks have more servers assigned to the *Cold* zone than the outer-most racks (some may even not have any servers assigned to the *Cold* zone). This policy has the most potential to save cooling energy costs in T^*GreenHDFS. For example, if k is as large as 30%, then with an inlet temperature distribution as shown in the Figure 2, all the servers with temperature higher than 21 will get assigned to the *Cold* zone leading to significant alleviation of the hot spots in the cluster. However, the cooling energy costs savings may come with some performance tradeoffs as illustrated below.

Big Data analytics framework such as MapReduce uses rack-awareness while writing data onto the cluster as intra-rack bandwidth is higher than inter-rack bandwidth. Two replicas of each block are written on servers in the same rack to take advantage of the intra-rack bandwidth for reduction in the writing time. If a rack (e.g., one of the central racks), has a large number of servers assigned to the *Cold* zone servers, then it would have less number of servers in the *Hot* zone and thereby, the replication pipeline at the time of the writes may not be able to find servers on the same rack to write the replica. This may result in an increase in the write latency.

The File Migration Policy discussed in Section 3.2.4, utilizes rack-awareness while migrating data from the *Hot* to the *Cold* zone, to take advantage of the higher intra-rack network bandwidth compared to the lower inter-rack network bandwidth; and, aims to migrate cold data from a *Hot* zone server to a *Cold* zone server residing on the same rack. If there are no servers assigned to the *Cold* zone in some racks, then the File Migration Policy has to resort to inter-rack migration, thus taking longer to migrate the data.

Rack-Level Zone Partitioning In this scheme, T^*GreenHDFS partitions servers in each rack between the *Hot*, and the *Cold* zone. The $k/(n_{Rack} * n_{ServersPerRack})$ most cooling-inefficient servers per rack are assigned to the *Cold*, and the rest of the servers in the rack to the *Hot* zone. This partitioning results in a uniform allocation of servers per rack to the zones. The Rack-Level partitioning has a performance advantage over the Cluster-Level Zone Partitioning as it offers better migration bandwidth and write performance. However, it may result in lower cooling energy savings as it may not be able to assign all the hot spots to the *Cold* zone; and, thereby may not be able to alleviate all the hot spots in the cluster.

Hardware Provisioning Since, cyber-controller places only cold data (i.e., very low or negligible data access-profile and computations, and thereby, very low thermal-profile) on the *Cold* zone servers, T^*GreenHDFS trades performance for aggressive energy savings. It uses low cost, low performance and low power processors in the servers in the *Cold* zone. Recently, several low-power processors have been introduced in

[1] floVENT simulates a cluster with great accuracy including the geometry, layout, and configuration of the compute equipment. floVENT has been extensively used and validated in several research papers in the past [6, 31, 37].

[2] The algorithm considers the distribution of the inlet temperatures in the cluster (e.g., the distribution in the Figure 2), and data and workload profiles in the cluster to come up with the number k.

the market. A class of Intel Atom introduced in 2010 called Z560 [20] is a single-core processor which consumes only 2.5W (0.01W when idle), has a clock speed of 2.13GHz, and costs $144. On the other hand, high power, high performance processors such a Quad-core Xeon 5400 are used in the servers in the *Hot* zone. A Quad-core Xeon 5400 consumes 80-150W of power while offering clock speeds ranging from 1.86-3.50GHz and its costs range from $209.00 - $1493.00 [21]. The *Hot* zone uses 8 DRAMs for high performance. The number of DRAMs is reduced from 8 to 2 in the *Cold* zone to further reduce the power consumption. The low power *Cold* zone servers can still be used for computations in situations where the *Hot* zone servers are not sufficient to absorb the entire workload such as in periods of heavy, peak demand.

3.2 Cyber Controller/Actuator

The cyber-controller shown in Figure 3 is the brain of T^*GreenHDFS. It gets file system events such as file create, read, and write from the file system clients and utilization, thermal, and capacity maps from the monitoring service running in the physical-system. Cyber-controller maintains predictive data models and uses the predicted file attributes in addition to the maps to guide its data placement decision and its policies. The predictive models, decision process, and policies are covered in the following sections.

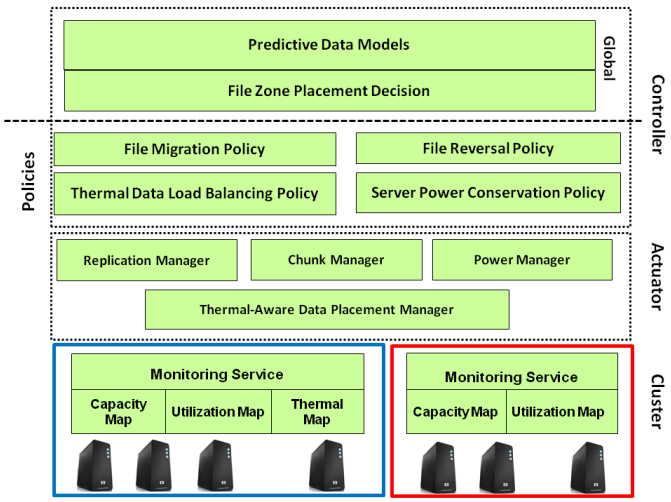

Figure 3: **Cyber-Side T^*GreenHDFS.**

3.2.1 Predictive Models

The cyber-controller maintains predictive models to predict file attributes from the absolute directory hierarchy of the file, at the time of file creation. The predictive data models are generated by using supervised learning on historical information present in the file system traces and metadata images. The predictor is based on our observations that there is a strong correlation between the directory hierarchy in which a file is organized and the file's attributes. Our earlier work details all aspects of the predictor such as training, testing and evaluation with real-world traces from production hadoop cluster at Yahoo! [23]. To figure out the statistical correlation between the directory hierarchies and file attributes, predictor uses Ridge Regression. Multiple Regression is a form of a supervised learning with input X (i.e., independent variables) and a response Y (i.e., dependent/response variable). The goal is to learn the correlation (regression) between X and Y. We treat subdirectories in the training data set, denoted as T, as independent input variables. The three file attributes: $Lifespan_{CLR}$ (i.e., the evolution lifetime of the file between the file create and the last access to file), file size and heat (i.e., access-profile of a file) are the dependent/response variables.

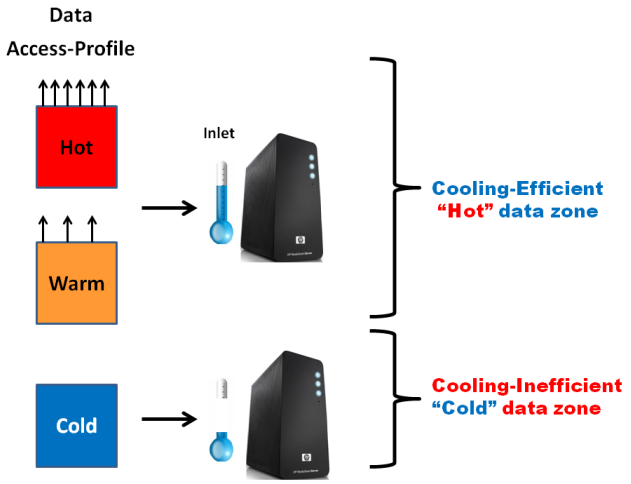

Figure 4: **Coarse-Grained File Zone Assignment.**

3.2.2 File Zone Placement Decision

When a file create event comes, cyber-controller uses the predictive data models to predict file's access-profile, size, and evolution lifespans. The cyber-controller decides the zone assignment of the file based on the predicted value of file access-profile as shown in Figure 4. Cyber-controller makes a coarse-grained decision about the file mapping to the zones. Files with predicted high, medium, and low access-profile are marked as candidates for the *Hot* zone. Files with very low access-profiles are marked as candidates for the aggressively power-managed *Cold* zone. Majority of the computations naturally happen in the *Hot* zone because of data-locality; thereby, *Hot* zone consumes significant energy. It becomes imperative to efficiently cool such servers to prevent risk of damage to the server components and system reliability. By using cooling-efficient servers in the *Hot* zone as discussed in Section 3.1.2, T^*GreenHDFS results in a lower overall thermal-profile.

On the other hand, data placed in the *Cold* zone is almost cold (i.e., dormant); data-locality results in very few (almost none) computations on these servers and hence, the *Cold* zone servers generate significantly less heat and do not need much cooling energy. Such data is a great fit for the most cooling-

inefficient servers in the data centers. Placing cold data on the cooling-inefficient servers ensures that the exhaust temperature of these servers doesn't increase and thereby, results in a lower and more uniform thermal-profile in the cluster. This results in lower cooling energy costs. On the other hand, placing data with high access-profile, and thereby, high thermal-profile would have been disastrous for the cooling-inefficient servers as such data would end up increasing the exhaust temperatures of these servers to high ranges; such hot spots would result in an overall increase in the temperature of the cluster and thereby, higher cooling costs. Thus, predictive file assignment to zones implicitly yields proactive cooling energy savings.

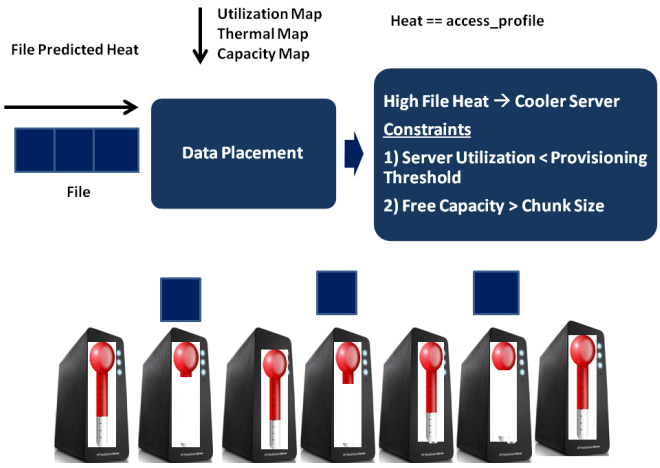

Figure 5: **Fine-Grained Thermal-Aware Data Placement.**

3.2.3 Thermal-Aware Data Placement Manager

Thermal-Aware Data Placement considers the problem of assigning n file chunks $f_1, f_2, \ldots, f_n$ among m servers in a fine-grained, thermal-, performance- and capacity-aware manner as shown in Figure 5. We assume that $m \geq n$. Thermal-Aware Data Placement matches the predicted data semantics (cyber-side) with the thermal-profile (physical-side) of the servers. A predicted high access-profile of a file gives an indication of high computation-profile of the file; and thereby, a high thermal-profile of the file. T^*GreenHDFS consults the run-time thermal map, heat map, and capacity map of the servers in the cluster which were captured by the monitoring service before placing the chunks on the servers. T^*GreenHDFS places the chunks of the files with predicted high access-profile on servers with lower run-time exhaust temperatures in order to ensure more uniform and lower thermal-profile in the data center to lower the cooling energy costs.[3] If the predicted access-profile is medium, the file is placed on servers that are relatively warm. If the access-profile is relatively low, the files are placed on the warmest servers in the *Hot* zone. Thermal-aware data placement is subject to two constraints: 1) the destination servers should have enough free capacity available to host the incoming block; and 2) the utilization on the destination servers should be less than the provisioning threshold to ensure that the performance of the server is not impacted adversely by the addition of the new file chunks and the associated computations. The thermal-aware, proactive data placement of T^*GreenHDFS allows more uniform thermal-profile, lower overall cluster temperature, and thereby, lower cooling costs.[4]

3.2.4 File Migration Policy

The File Migration Policy identifies the files that have become dormant (i.e., are no longer accessed) so that they can be migrated to the *Cold* zone. T^*GreenHDFS uses the predictive models to predict the $Lifespan_{CLR}$ (lifespan between the file create and last file access) of a file at the file creation time and migrates files in a proactive, self-adaptive, and per-file fine-grained manner at the end of file's $Lifespan_{CLR}$. The details on File Migration Policy and evaluations such as data moved per day, performance impact of the movement are covered in our work [23, 24].

3.2.5 Server Power Conserver Policy

Cyber-controller invokes an energy saving policy called *Server Power Conserver*. The policy monitors the data access activity to the servers in the *Cold* data zone and scales-down a server if it hasn't been accessed in the last n threshold number of days. The disks are transitioned to the sleep power mode, processors are set to ACPI S4 "Sleep" state defined by the ACPI standard, and DRAM is put in self-refresh operating mode. T^*GreenHDFS relies on Wake-on-LAN support in NICs to transition servers back to active power mode upon a future access or event such as bit-rot checker. Details of scale-down, its mandates, and performance considerations are covered in our earlier work [23, 24].

3.3 Physical Controller/Actuator

The physical-controller monitors the overall temperature of the physical-system and with the help of the physical-actuator, controls the CRAC's outlet temperature based on the overall temperature of the physical system. The following equations elaborate the logic behind the physical-controller.

The efficiency of the CRACs is characterized by their coefficient of performance (COP). COP is defined as the ratio of the amount of heat Q removed by the cooling device to the energy W consumed by the cooling device. Thus, work required to remove heat is inversely proportional to the COP.

$$COP(T_{out}) = \frac{Q}{W} \quad (1)$$

[3] Placing such files on servers whose exhaust temperature is already high, would result in an even higher exhaust temperature; indirectly resulting in higher overall cluster temperature and increased cooling energy costs.

[4] The Cyber Actuator has several per-zone components such as chunk manager, and replication manager which decide the number of chunks and replicas for the file. In this paper, T^*GreenHDFS assumes a chunk size of 128MB and 3-way replication as is a norm in the production clusters.

Q is calculated as shown below [33]:

$$Q = m * C_p * (T_{in} - T_{out}) \quad (2)$$

A typical COP model obtained from a Liebert CRAC unit [6]:

$$COP(T_{out}) = \left(0.0068T_{out}^2 + 0.0008T_{out} + 0.458\right); \quad (3)$$

Where, T_{in} is the temperature of the hot air that needs to be cooled by the CRAC, T_{out} is the temperature of the cold air supplied by the CRAC, m is mass flow rate, and C_p is the specific heat.

The physical-actuator controls the CRAC's air supply temperature T_{out} according to the overall temperature observed in the physical system. If the overall temperature in the physical system has cooled down below a threshold temperature, the physical-actuator increases the outlet temperature T_{out} of the cooling subsystem while ensuring that the inlet temperatures of the servers remain below the redline temperature (as that would pose a risk to the reliability). Increasing T_{out} in turn increases the efficiency (COP) of the CRAC as shown in Equation 3 and results in lower cooling energy costs. The physical-actuator reduces T_{out} if the overall temperature of the cluster becomes high again for any reason.

4 Evaluation

We use Mentor Graphic's floVENT, a computational Fluid Dynamics (CFD) simulator [1].[5] The cluster under evaluation has four rows of fourteen industry standard 47U racks arranged in a typical cold aisle and hot aisle layout [38]. The racks and perforated floor tiles are installed on the raised floor and CRAC delivers cold air under the elevated floor. The cool air enters the racks from the front (where server inlets are located), picks up heat while flowing through these racks, and exits from the rear of the racks (where server exhausts are located). The heated air is extracted back to the CRAC intakes that are positioned above the hot aisles. The cold supply air rises from raised floor plenum through vent tiles, and hot exhaust air returns to the CRAC through ceiling vent tiles. Each rack contains 46, 1U servers for a total of 2576 servers.[6] Each 1U server consumes 150W at peak utilization and 90W at idle utilization. The cold air is supplied by two 180KW computer room air conditioner units (CRACs), with the flow rate of 30 m^3/s. The CRACs' supply temperature is fixed at 16^{o}C.

We consider different configurations in our simulation to isolate the advantages and contribution of the various features of T^*GreenHDFS to the overall cooling energy costs reduction. The configurations are listed in the Table 3.3. To simulate HDFS, we do a floVENT simulation where each server in the cluster is loaded at 30%-50% utilization.[7] Given, the non-energy-proportional nature of the servers, 30%-50% utilization draws 82% of the peak power [17]. To simulate T^*GreenHDFS, the *Hot* zone servers are shown to be at 100% utilization which is very conservative in nature (in reality servers will be between 50%-70% utilization). This is chosen to compare the worst-case energy profile of T^*GreenHDFS with the best-case energy profile of the baseline HDFS.

We evaluate with synthetically generated traces that have same characteristics as the one-month long real-world HDFS traces generated by a production (2600 servers, 34 million files, 5 Petabytes) Hadoop cluster at Yahoo! that we had used in our earlier work [24]. We focus our analysis on the biggest (60% of the used storage capacity) and most important data set (clickstream) in the production cluster. Log processing is one of the most popular use cases of data-intensive computing in the web 2.0 Internet services companies such as Facebook, Google, and Yahoo! [10]. These companies rely on clickstream processing [15], an example of log processing, to calculate the web-based advertising revenues, and derive user interest models and predictions. For this, daily huge web logs are analyzed in the production environments [40]. Next, we present the evaluation results.

4.1 Cooling Energy Costs

As shown in Figure 6, T^*GreenHDFS_T_Placement results in 47%, T^*GreenHDFS_Overall_30 37%, and T_Placement_Only 29% reduction in cooling energy costs compared to baseline *HDFS*. T^*GreenHDFS_Overall_30 owes its savings to thermal-aware zone partitioning whereby the inherently most cooling-inefficient servers in the cluster are assigned to the *Cold* data zone; the low computation profile (hence, thermal-profile) of the cold data placed on these servers thwarts development of thermal hot spots (i.e., servers with high exhaust temperatures). Since, the hot spots are the primary cause of high overall temperature in the cluster and high cooling costs, T^*GreenHDFS results in lower overall temperature and hence, lower cooling energy costs. T^*GreenHDFS_T_Placement has even lower cooling energy costs than T^*GreenHDFS_Overall_30 because it additionally does predictive, thermal-aware data placement in the *Hot* data zone. As seen in Figure 2, there is significant variation in the inlet and run-time exhaust temperatures in the cluster, and thermal-aware data placement places data in such a way that the computation-profile of the data is inversely proportional to the temperature of the server. T_Placement_Only doesn't do any data migration, zoning, or scale-down; instead, it does predictive, thermal-aware data placement covered in the Section 3.2.3 in the entire cluster. The high savings possible with T_Placement_Only are encouraging and provide an alternative mechanism to save cooling energy costs if zoning, data migration and scale-down are not an option for an enterprise. In Figure 7, we show normalized cooling savings with or without scale-down of the *Cold* zone servers. Even without scale-down T^*GreenHDFS results in cooling energy costs savings courtesy of its thermal-awareness. However, zoning, data migration and scale-down are needed for saving server operating energy costs; which we had shown to be 26% in GreenHDFS [24].

[5] floVENT has been extensively used and validated in several research papers in the past [6, 31, 37].

[6] Same size as the large-scale production Hadoop cluster that we had used in our earlier papers [24].

[7] Studies have shown that the typical utilization of the servers in the Big Data Analytics compute cluster is 30%-50% [17].

Table 1: **Evaluation Configurations**

Configuration	Options	Explanation
HDFS		Baseline. Cluster with no energy- or thermal-awareness or data-classification driven data placement.
T^*GreenHDFS_Rack		T^*GreenHDFS uses thermal-aware rack-aware zone partitioning covered in Section 3.1.2 to create thermal-aware Hot and Cold zones. T^*GreenHDFS does data migration, and server scale-down. There is no thermal-aware fine-grained data placement done in Hot zone in this scenario.
	T^*GreenHDFS_Rack_10	10% cluster servers in Cold zone and 90% in Hot zone.
	T^*GreenHDFS_Rack_20	20% cluster servers in Cold zone and 80% in Hot zone.
	T^*GreenHDFS_Rack_30	30% cluster servers in Cold zone and 70% in Hot zone.
T^*GreenHDFS_Overall		T^*GreenHDFS uses cluster-level zone partitioning as covered in Section 3.1.2 to create thermal-aware Hot and Cold zones . T^*GreenHDFS does data migration, and server scale-down. There is no thermal-aware fine-grained data placement done in Hot zone in this scenario.
	T^*GreenHDFS_Overall_10	10% cluster servers in Cold zone and 90% in Hot zone.
	T^*GreenHDFS_Overall_20	20% cluster servers in Cold zone and 80% in Hot zone.
	T^*GreenHDFS_Overall_30	30% cluster servers in Cold zone and 70% in Hot zone.
T^*GreenHDFS_NoScaleDown		Same as T^*GreenHDFS_Overall_30, but doesn't do any server scale-down in the Cold zone
T^*GreenHDFS_T _Placement		Same as T^*GreenHDFS_Overall_30. In addition, does thermal-aware fine-grained data placement in the Hot zone as covered in Section 3.2.3.
T_Placement_Only		T^*GreenHDFS doesn't divide the cluster into zones, and doesn't do any data migration or server scale-down. T^*GreenHDFS only uses thermal-aware, predictive, fine-grained data placement cluster-wide.

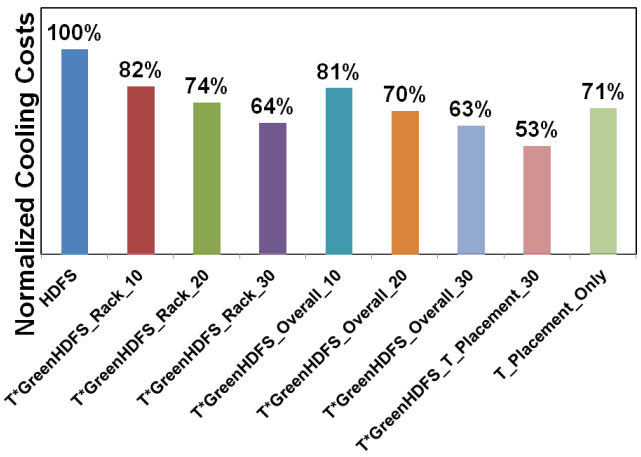

Figure 6: **The Cooling Energy Costs with Different Configurations of Thermal-Aware T^*GreenHDFS Normalized to Baseline *HDFS*.**

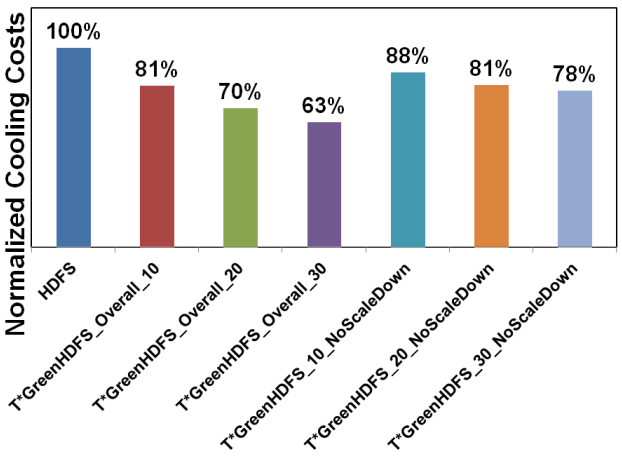

Figure 7: **The Cooling Energy Costs with Thermal-Aware T^*GreenHDFS with and without Scale-Down Normalized to Baseline *HDFS*.**

Next, we compare the cooling costs reduction possible in T^*GreenHDFS with different Zone Partitioning schemes covered in Sections 3.1.2 and 3.1.2. As shown in Figure 6, the cooling costs reduction is very similar with Rack-Level Zone Partitioning and Cluster-Level Zone Partitioning. Since, Rack-Level Zone Partitioning has performance advantages over Cluster-Level Zone Partitioning with respect to data migration and data writes; it can be used without any cooling costs tradeoffs. In each of the Zone Partitioning schemes we consider a different split (i.e., k) between the *Hot* data zone and the *Cold* data zone. The purpose of this experimentation is to show that T^*GreenHDFS is capable of saving both server and cooling energy costs even if as low as 10% servers are assigned to the *Cold* data zone (only candidates for scale-down in the cluster). As shown in Figure 6, T^*GreenHDFS_Rack_10 and T^*GreenHDFS_Overall_10 are both capable of saving 9-10% cooling energy costs and 12% server energy costs.

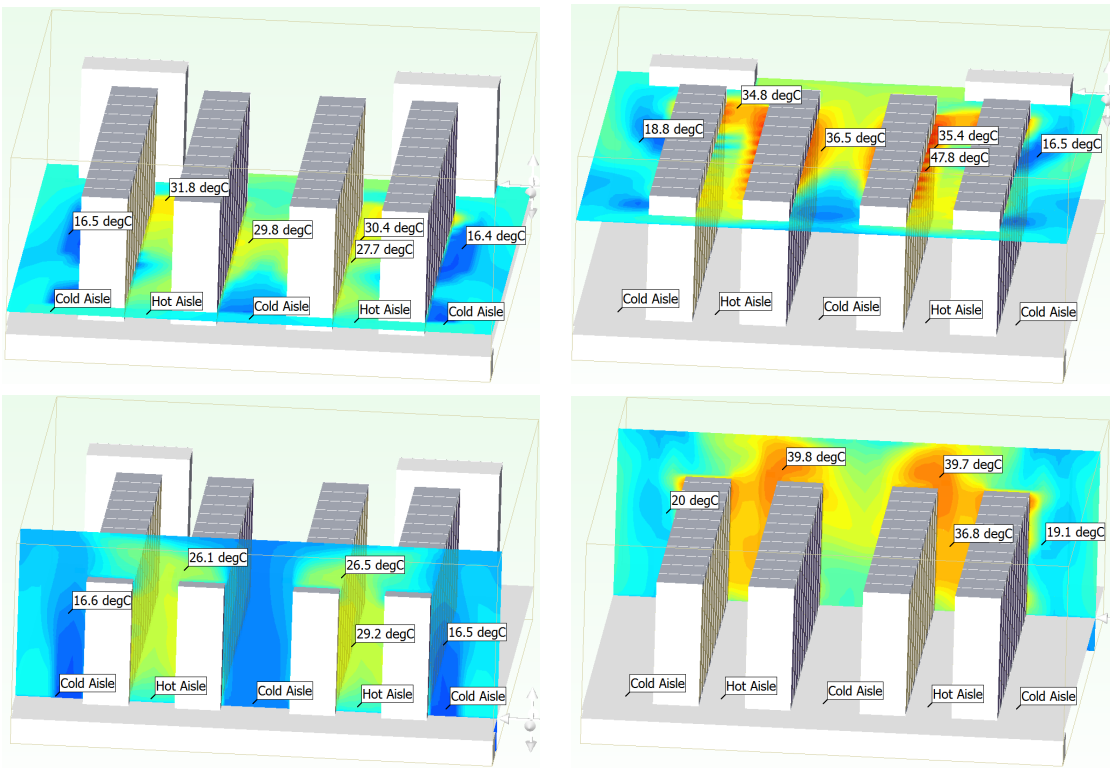

Figure 9: **Thermal Contour Plots of Baseline HDFS (No Thermal Management)**

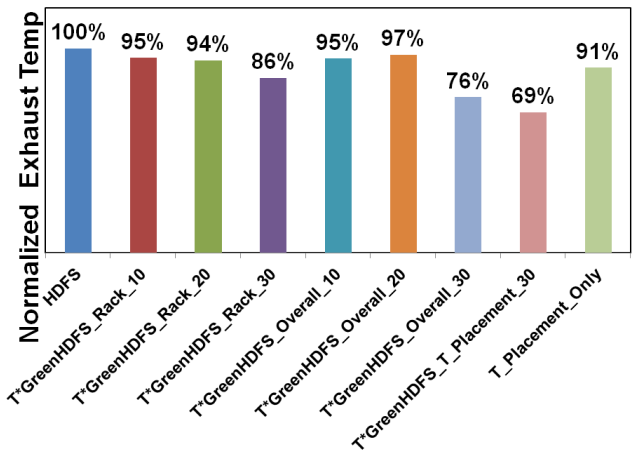

Figure 8: **Maximum Server Exhaust Temperatures Normalized to Baseline *HDFS*.**

4.2 Server Exhaust Temperature

Figure 8 shows the maximum exhaust temperature of the servers, normalized with respect to the baseline *HDFS*, under different configurations. T^*GreenHDFS_T_Placement results in a 31% reduction in the maximum exhaust temperature in the servers. T^*GreenHDFS can reduce the cooling costs additionally by increasing the temperature T_{out} of the air supplied by the CRAC (of course, while ensuring no server exceeds the redline temperature. This increases the COP of the CRAC as shown in Equation (3), and allows CRAC to operate at higher efficiency. For example, operating the CRACs at 5^oC higher supply temperature of 20^oC results in increasing COP from 5 to 6 as shown in Equation 3. This increase in the COP results in an additional 13% cooling costs savings. All the T^*GreenHDFS configurations result in a reduction in the maximum exhaust temperature in the cluster, and hence CRAC can be operated at proportionately higher temperature for all the schemes leading to additional cooling costs savings.

4.3 Thermal Profile

To evaluate the impact of thermal-aware zoning and thermal-aware, proactive data placement in T^*GreenHDFS, we compare the exhaust and inlet temperatures of the servers and the thermal contour plots under different T^*GreenHDFS configurations. Figure 12 shows the thermal contour plots of T^*GreenHDFS_T_Placement at different planes in the cluster. The color coding spectrum in the figures ranges from Blue (16^oC) to Red (50^oC) color. The T^*GreenHDFS_T_Placement thermal profiles are much more uniform, have lower temperatures and less hot spots than the baseline *HDFS* thermal plots shown in Figure 12.

Figure 11 shows the exhaust and inlet temperatures and overall, T^*GreenHDFS_T_Placement has the lowest and most uniform temperatures across the cluster and is followed by T^*GreenHDFS_Overall_30. Lower inlet temperatures indicate a lowering of hot air recirculation in the cluster, which in turn reduces thermal hot spots in the cluster. The uniform thermal profile is a result of the fine-grained, proactive,

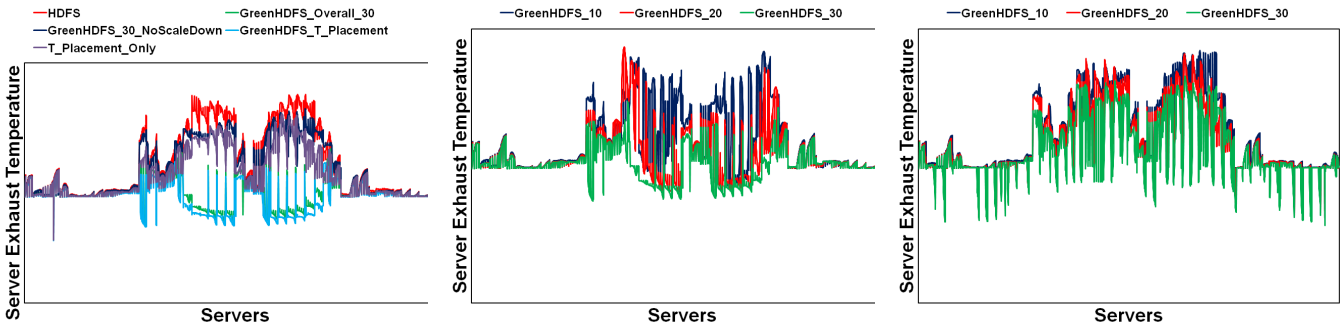

Figure 10: **The Exhaust Temperatures of the Servers in T^*GreenHDFS a) Main Configurations Compared with *HDFS*, b) Cluster-Level Zone Partitioning, and c) Rack-Level Zone Partitioning. Temperatures are the Lowest and the Most Uniform with T^*GreenHDFS_T_Placement.**

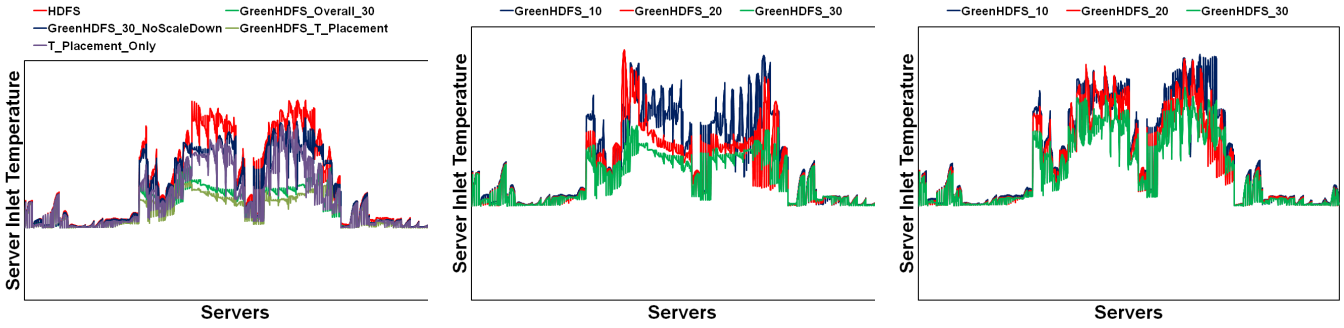

Figure 11: **The Inlet Temperatures of the Servers in T^*GreenHDFS a) Main Configurations Compared with *HDFS*, b) Cluster-Level Zone Partitioning, and c) Rack-Level Zone Partitioning. The Temperatures are the Lowest and the Most Uniform with T^*GreenHDFS_T_Placement.**

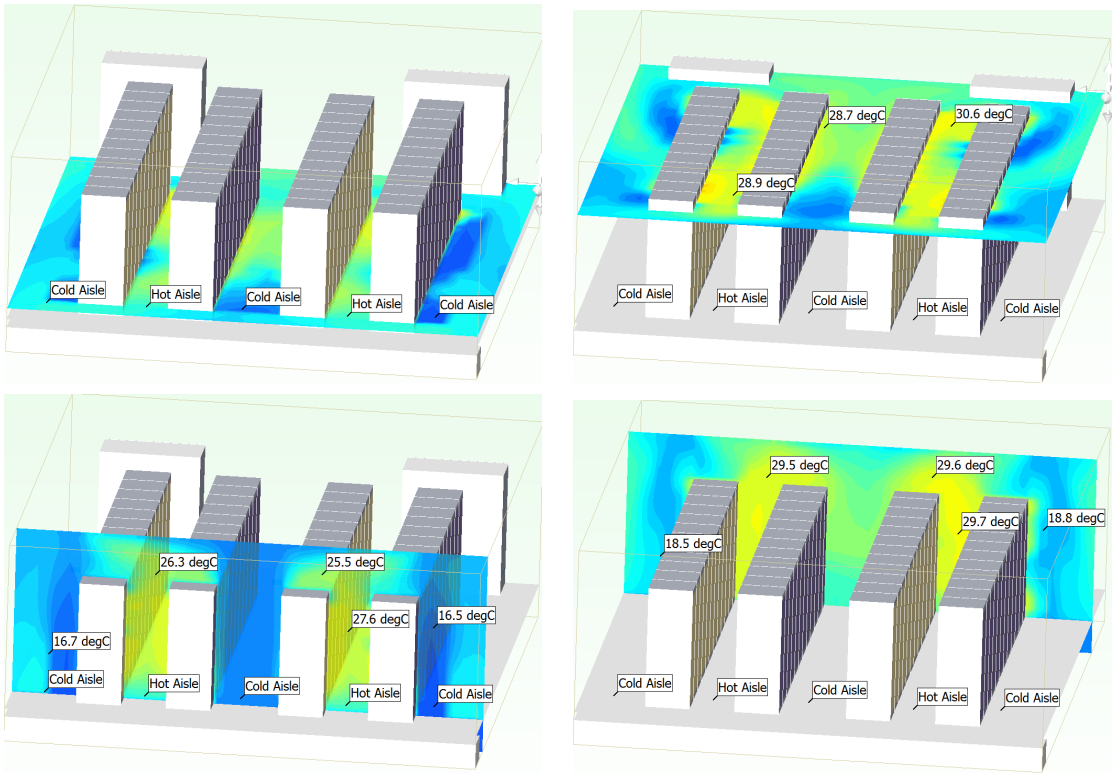

Figure 12: **Thermal Contour Plots with *T^*GreenHDFS_T_Placement*.**

thermal-aware data placement in T^*GreenHDFS. In summary, T^*GreenHDFS is able to reduce the cooling costs by lowering the overall temperature in the cluster, making the thermal-profile much more uniform, and by reducing the hot air recirculation.

5 Related Work

Cooling strategies can be broadly divided into server-level, ensemble-level, and data-center-level strategies. Server-level examples include active server fan tuning to cool down the servers [29]. Cohen et al. propose control strategies via DVFS to enforce constraints on the chip temperature and on the workload execution [22]. At the ensemble level, Niraj et. al. rely on workload migration and location-dependent cooling-efficiency of the fans to manage the power and thermal characteristics of the ensemble.

There is significant research on reducing cooling energy costs at the data-center-level [11, 34, 33, 8, 7, 43, 39, 12, 35]. The research on cooling-efficient data center layouts, models and server and rack designs [38, 36, 32] is orthogonal to T^*GreenHDFS. The run-time strategies are mostly *task-centric* in nature [31, 37, 6, 41, 4, 39, 35, 30]. For example, Moore et. al. [31] provide a mechanism to do temperature-aware workload placement. Sharma et. al. [37] present a framework for thermal load balancing whereby they show how an asymmetric, thermal-aware workload placement and migration can result in uniform temperature distribution in the data center. Bash et. al. [6] attempt to place heavy computational workloads on servers in cooling-inefficient locations in the data center. Sarood et. al. do thermal-aware load balancing [35]. Parolini et. al. and Tang et. al. present a cyber-physical systems approach for data center modeling and control for energy-efficiency which is again task-centric in nature [39, 28].

Big Data Analytics cloud presents a challenge to the task-centric techniques. In Big Data analytics compute cloud model, data-locality is an important consideration for network-efficient, high performance computing; computations are sent to the data as sending data to the computations isn't network-efficient and separation of storage and compute nodes isn't cost-efficient. Data-locality consideration and significant server state *limit* thermal-aware task placement and task migration based cooling techniques. Given the explosion in Big Data, data needs to become a first-class object in computing and the computing paradigms need to change accordingly. T^*GreenHDFS takes a data-centric approach and does proactive, thermal-aware data placement which in turn leads to thermal-aware computation placement.

Recent research on scale-down in MapReduce GFS and HDFS managed clusters seeks to exploit the replication feature of these file systems and proposes energy-aware replica placement techniques for server scale-down [5]. Lang and Patel propose an "All-In" strategy (AIS) for scale-down in MapReduce clusters [26]. These techniques are not thermal-aware and focus only on the computing energy costs savings. T^*GreenHDFS takes a different thermal-aware, data-centric, data-classification driven approach to scale-down servers and results in both computing and cooling energy costs savings.

Vasudevan et. al. [42] have proposed data-intensive clusters built with low power, lower performance (Wimpy) servers that aim to reduce the peak power consumption of the cluster. Lang et. al. [27] point out that for more complex workloads such clusters will result in a more expensive and less performant solution. T^*GreenHDFS uses low power, low performance servers only in a small subset of the cluster called the *Cold* zone where performance is not an important criterion.

6 Conclusion

Massive proliferation of data-intensive cloud computing clusters mandates a reduction in the overall energy costs. We present a thermal-aware cloud file system T^*GreenHDFS which takes a data-centric approach to reduce cooling energy costs in Big Data analytics cloud. On the physical-side, T^*GreenHDFS is cognizant of the thermal-profile of the servers in the cluster. On the cyber-side, T^*GreenHDFS is aware of the data semantics. Based on this knowledge, and coupled with its predictive data models, T^*GreenHDFS does proactive, thermal-aware data placement and thermal-aware server partitioning. T^*GreenHDFS results in more uniform thermal-profile, and lower overall temperature in the cluster. Evaluation results show upto a 59% reduction in the cooling energy costs.

References

[1] floVENT: A Computational Fluid Dynamics Simulator. `http://www.mentor.com/products/mechanical/products/flovent`.

[2] The Green Grid Data Center Power Efficiency Metrics: PUE and DCiE, 2007.

[3] Cost of Power in Large-Scale Data Centers. `http://perspectives.mvdirona.com`, November, 2008.

[4] ABBASI, Z., VARSAMOPOULOS, G., AND GUPTA, S. K. S. Thermal aware server provisioning and workload distribution for internet data centers. In *HPDC '10: Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing* (New York, NY, USA, 2010), ACM, pp. 130–141.

[5] AMUR, H., CIPAR, J., GUPTA, V., GANGER, G. R., KOZUCH, M. A., AND SCHWAN, K. Robust and Flexible Power-Proportional Storage. In *Proceedings of the 1st ACM Symposium on Cloud Computing* (New York, NY, USA, 2010), SoCC'10, ACM, pp. 217–228.

[6] BASH, C., AND FORMAN, G. Cool Job Allocation: Measuring the Power Savings of Placing Jobs at Cooling-Efficient Locations in the Data Center. In *USENIX Annual Technical Conference* (2007), ATC'07, pp. 363–368.

[7] BASH, C., PATEL, C., AND SHARMA, R. Dynamic thermal management of air cooled data centers. In *Thermal and Thermomechanical Phenomena in Electronics Systems, 2006. ITHERM '06. The Tenth Intersociety Conference on* (30 2006-june 2 2006), pp. 8 pp. –452.

[8] BEITELMAL, M. H., AND PATEL, C. D. Model-based approach for optimizing a data center centralized cooling system. Tech. rep., Hewlett Packard, 2006.

[9] BELL, G., HEY, T., AND SZALAY, A. Beyond the data deluge.

[10] BLANAS, S., PATEL, J. M., ERCEGOVAC, V., RAO, J., SHEKITA, E. J., AND TIAN, Y. A Comparison of Join Algorithms for Log Processing in MapReduce. In *Proceedings of the 2010 International Conference on Management of Data* (New York, NY, USA, 2010), SIGMOD '10, ACM, pp. 975–986.

[11] BREEN, T., WALSH, E., PUNCH, J., SHAH, A., AND BASH, C. From chip to cooling tower data center modeling: Part i influence of server inlet temperature and temperature rise across cabinet. In *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2010 12th IEEE Intersociety Conference on* (june 2010), pp. 1 –10.

[12] CHEN, Y., GMACH, D., HYSER, C., WANG, Z., BASH, C., HOOVER, C., AND SINGHAL, S. Integrated management of application performance, power and cooling in data centers. In *Network Operations and Management Symposium (NOMS), 2010 IEEE* (april 2010), pp. 615 – 622.

[13] Data-intensive computing. `http://research.yahoo.com/files/BryantDISC.pdf`.

[14] DEAN, J., GHEMAWAT, S., AND INC, G. MapReduce: Simplified Data Processing on Large Clusters. In *Proceedings of the 6th Conference on Symposium on Operating Systems Design and Implementation* (2004), OSDI'04, USENIX Association.

[15] DEPARTMENT, P. C., CHATTERJEE, P., HOFFMAN, D. L., AND NOVAK, T. P. Modeling the Clickstream: Implications for Web-Based Advertising Efforts. *Marketing Science 22* (2000), 520–541.

[16] EPA. EPA Report on Server and Data Center Energy Efficiency. Tech. rep., U.S. Environmental Protection Agency, 2007.

[17] FAN, X., WEBER, W.-D., AND BARROSO, L. A. Power Provisioning for a Warehouse-Sized Computer. In *ISCA '07: Proceedings of the 34th Annual International Symposium on Computer Architecture* (New York, NY, USA, 2007), ACM, pp. 13–23.

[18] GREENBERG, S., MILLS, E., TSCHUDI, B., RUMSEY, P., AND MYATT, B. Best Practices for Data Centers: Results from Benchmarking 22 Data Centers. In *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings* (2006).

[19] HOELZLE, U., AND BARROSO, L. *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*. Morgan and Claypool Publishers, May 29, 2009.

[20] INTEL. Intel Atom Processor Z560. `http://ark.intel.com/Product.aspx?id=49669\&processor=Z560\&spec-codes=SLH63`.

[21] INTEL. Quad-core intel xeon processor 5400 series price. `http://ark.intel.com/ProductCollection.aspx?series=33905`, 2008.

[22] JIAN-HUI, Z., AND CHUN-XIN, Y. Design and simulation of the cpu fan and heat sinks. *Components and Packaging Technologies, IEEE Transactions on 31*, 4 (dec. 2008), 890 –903.

[23] KAUSHIK, R. T., ABDELZAHER, T., AND NAHRSTEDT, K. Predictive Data and Energy Management in GreenHDFS. In *Proceedings of International Conference on Green Computing* (2011), IGCC, IEEE.

[24] KAUSHIK, R. T., BHANDARKAR, M., AND NAHRSTEDT, K. Evaluation and Analysis of GreenHDFS: A Self-Adaptive, Energy-Conserving Variant of the Hadoop Distributed File System. In *Proceedings of the 2nd IEEE International Conference on Cloud Computing Technology and Science* (2010), CloudCom, IEEE.

[25] KONSTANTIN, S., KUANG, H., RADIA, S., AND CHANSLER, R. The Hadoop Distributed File System. *MSST* (2010).

[26] LANG, W., AND PATEL, J. M. Energy Management for MapReduce Clusters. *Proceedings VLDB Endow. 3* (September 2010), 129–139.

[27] LANG, W., PATEL, J. M., AND SHANKAR, S. Wimpy Node Clusters: What about Non-Wimpy Workloads? In *Proceedings of the Sixth International Workshop on Data Management on New Hardware* (New York, NY, USA, 2010), DaMoN '10, ACM, pp. 47–55.

[28] LUCA PAROLINI, BRUNO SINOPOLI, B. H. K., AND WANG, Z. A Cyber-Physical-System Approach to Data Center Modeling and Control for Energy Efficiency. In *Proceedings of the IEEE, Special Issue on Cyber-Physical Systems* (December 2011).

[29] MAHAJAN, R., PIN CHIU, C., AND CHRYSLER, G. Cooling a microprocessor chip. *Proceedings of the IEEE 94*, 8 (aug. 2006), 1476 –1486.

[30] MOORE, J., CHASE, J., AND RANGANATHAN, P. Weatherman: Automated, online and predictive thermal mapping and management for data centers. In *Autonomic Computing, 2006. ICAC '06. IEEE International Conference on* (june 2006), pp. 155 – 164.

[31] MOORE, J., CHASE, J., RANGANATHAN, P., AND SHARMA, R. Making Scheduling "Cool": Temperature-Aware Workload Placement in Data Centers. In *Proceedings of the Annual Conference on USENIX Annual Technical Conference* (Berkeley, CA, USA, 2005), ATEC '05, USENIX Association, pp. 5–5.

[32] PATEL, C., BASH, E., SHARMA, R., AND BEITELMAL, M. Smart Cooling of Data Centers. In *Proceedings of PacificRim/ASME International Electronics Packaging Technical Conference and Exhibition* (2003), IPACK'03.

[33] PATEL, C., SHARMA, R., BASH, C. E., AND BEITELMAL, A. Thermal Considerations in Cooling Large Scale High Compute Density Data Centers. In *Eight Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems* (2002), ITherm'02, pp. 767–776.

[34] PATEL, C. D., BASH, C. E., BELADY, C., STAHL, L., AND SULLIVAN, D. Computational Fluid Dynamics Modeling of High Compute Density Data Centers to Assure System Inlet Air Specifications.

[35] SAROOD, O., AND KALE, L. V. A 'cool' load balancer for parallel applications. In *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis* (New York, NY, USA, 2011), SC '11, ACM, pp. 21:1–21:11.

[36] SCHMIDT, R. R., KARKI, K. C., KELKAR, K. M., RADMEHR, A., AND PATANKAR, S. V. Measurements and predictions of the flow distribution through perforated tiles in raised-floor data centers. In *Proceedings of The Pacific Rim/ASME International Electronic Packaging, IPACK* (2001).

[37] SHARMA, R. K., BASH, C. E., PATEL, C. D., FRIEDRICH, R. J., AND CHASE, J. S. Balance of Power: Dynamic Thermal Management for Internet Data Centers. *IEEE Internet Computing 9* (2005), 42–49.

[38] SULLIVAN, R. Alternating Cold and Hot Aisles Provides more Reliable Cooling for Server Farms.

[39] TANG, Q., GUPTA, S., AND VARSAMOPOULOS, G. Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach. *Parallel and Distributed Systems, IEEE Transactions on 19*, 11 (nov. 2008), 1458 – 1472.

[40] THUSOO, A., SHAO, Z., ANTHONY, S., BORTHAKUR, D., JAIN, N., SEN SARMA, J., MURTHY, R., AND LIU, H. Data Warehousing and Analytics Infrastructure at Facebook. In *Proceedings of the 2010 International Conference on Management of Data* (New York, NY, USA, 2010), SIGMOD '10, ACM, pp. 1013–1020.

[41] TOLIA, N., WANG, Z., MARWAH, M., BASH, C., RANGANATHAN, P., AND ZHU, X. Delivering Energy Proportionality with Non Energy-Proportional Systems: Optimizing the Ensemble. In *Proceedings of the 2008 Conference on Power Aware Computing and Systems* (Berkeley, CA, USA, 2008), HotPower'08, USENIX Association, pp. 2–2.

[42] VASUDEVAN, V., ANDERSEN, D., KAMINSKY, M., TAN, L., FRANKLIN, J., AND MORARU, I. Energy-Efficient Cluster Computing with FAWN: Workloads and Implications. In *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking* (New York, NY, USA, 2010), e-Energy '10, ACM, pp. 195–204.

[43] WANG, Z., TOLIA, N., AND BASH, C. Opportunities and challenges to unify workload, power, and cooling management in data centers. In *Proceedings of the Fifth International Workshop on Feedback Control Implementation and Design in Computing Systems and Networks* (New York, NY, USA, 2010), FeBiD '10, ACM, pp. 1–6.

[44] WHITE, T. *Hadoop: The Definitive Guide*. O'Reilly Media, May, 2009.